A Smartphone Assistance for Blind & Visually Impaired using Deep Learning: A Glimpse at Challenges

Jayashankar Chaluvaraj

Dissertation submitted in part fulfilment of the requirements for the master's degree in Data Analytics at Dublin Business School

Supervisor: Basel Magableh

January 2020

Declaration

I declare that this dissertation that I have submitted to Dublin Business School for the award of Master Science in Data Analytics is the result of my ideas and where others' ideas or words have been included, I have adequately cited and referenced the sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Signed: Jayashankar Chaluvaraj

Dublin Business School

Student Number: 10506661

Date: 06 – 01 – 2020

Acknowledgements

I wish to express my sincere gratitude to Mr. Basel Magableh, Supervisor for providing valuable feedback during the course of the dissertation.

I sincerely thank Mr. Naman Shah for their guidance and encouragement in all the technical ideas and motivated pursuing a carrier in Data Analytics.

Table of Content

Table of Figures

Abstract

The study explains the primary challenges in implementing the smart assist aid to the blind and visually impaired people to navigate easily and safely to a planned destination without depending on another person. Besides, tremendous efforts and resources are invested against autonomous car development using AI. However, none of these technologies are used toward people to navigate. The standard autonomous car principle is employed to overcome the obstacle, the typical challenges found are to detect the footpath lane pattern and train the model to detect the human turning angle in the curved street. The curvy road angle is perceived as a regression problem and a similar turn is categorized as a direction (Left & Right) to solve as a classification problem. Advanced algorithms are used to detect lane patterns and also illustrate the complexities and drawbacks associated with the process.

Chapter 1: Introduction

1.1 Background and aims

The Blindness and Vision Impairment people normally have challenges in stepping in an unfamiliar, strange and obscure place individually. The necessity for designing, developing and using as assistive devices become extremely crucial and fundamental for these sections of characters. To provide them autonomous navigation equipment with effective guidance on their move, various obstacles should be acknowledged. Firstly, the device has to locate a person and learn the surroundings just like a human to make an appropriate decision to guide. On the other hand, it has to identify the destination to reach the other end safely. Finally, calculate the shorted path from starting position to the destination, obstacle detection, traffic signal interpretation on the road.

In reality, blinded navigation is a huge issue faced by many people around the world. The World Health Organization estimates around at least 2.2 billion individuals have a vision impairment or blindness, at least 1 billion of whom have a vision impairment that could have been avoided or still needs to be addressed. Blindness incidence and demographics vary greatly from one part of the world to another. Around 1 billion individuals comprise those with mild to an extreme loss in vision or blindness due to unaddressed refractive error, as well as close vision impairment due to unaddressed presbyopia. Globally, uncorrected refractive defects and cataracts are the leading causes of vision impairment. Most people with vision impairment are over 50 years of age.

The common issue is encountered by these segment of people are while they try as simple as crossing the street, this is partially addressed by the traditional white cane guide dogs which are extremely good at helping blind people avoid obstacles. However, neither of them failed at providing visual aspects of the surrounding. The cane stick is not fully equipped to provide visual data of the enclosing environment around you, on the other hand, the guide dog has the color issue which fails to differentiate traffic signal and respond correspondingly. Some of these reasons discussed show difficulties in dealing with a simple solution like rely on cane stick or guide dog. Currently, the common resolution is eight help them cross the street or they use their hearing sense to capture the direction of traffic flow and then make decisions whether it is safe to cross the street. However, this also adequately does not address the problem. In case one, the dependency to rely on the other person increases for every action the blind person performs and in the other case, the waiting time is inevitable if the intersection is at the construction site, busy junction, rainy and noisy. The ability to predict and cross the road becomes though unless they can see visually and experience. Staying within the payment cross box boundaries and walking straight inline also become extremely challenging. Now we have a huge problem to solve and results in the potential requirement of a very advance solution to tackle this problem.

While analyzing this specific problem and completely unbelievable find is that the companies like Google, Tesla, NVIDIA and many more companies have put tones of effort and resources into helping cars navigate using sensors surrounding it. However, the same effort, investment, time and technology is not used equipment blind people navigate in our society who could benefit from the same technology. There has been a huge focus on self-driving cars and using object detection and lane detection to build their software. Similarly, the same principle is applied to help a blind to cross the road, walk in the park, climb the stairs, detect the obstacles and sense the environment using audio signals.

In theory, the self-driving car technology used to recognize traffic lights, lanes and obstacles is then implement applied to the smartphone device using the camera to assist blind people to recognize the surrounding environment using audio signals or by haptic feedback.

There are challenges associated with its development such as the accuracy the model is going to detect the object and guide the blind person and the other criteria are that it has to be fast and able to run in real-time, finally, the developed model has to run on low budget smart devices as well to cater the larger disable audience.

1.2 Thesis structure

In the research paper, the study begins with the literature review followed by methodologies and problems associated with it and finally end with future opportunities and conclusion.

Chapter 2: Literature review

2.1 Introduction

The review will enumerate, describe, summarize, objectively evaluate and clarify and overall includes a comprehensive summary of previous research on this topic with development strategies and the limitations. It will help distinguish the theoretical base for the research and help the researchers determine the nature of the investigation. It also acknowledges the work of previous researchers, and in so doing, assures the reader that the work has been well-conceived.

2.2 Report

Shraga Shoval et al. (1998) gives an overview of travel aid for blind people and to test this whole concept, they have a prototype device named NavBelt. This system can be used as a primary or secondary aid that comprises a portable computer, headphones and ultrasonic sensors. The obstacle and navigation techniques have also been developed by the mobile robots are being used as a primary source in this particular device after that the computer system uses a stereophonic imaging technique to process an ultrasonic sensor signal and transmitting information to the user's headphones. But the information received by the user is in the form of the acoustic picture of the whole environment suggesting user the direction of travel depending on operation mode also the signals are transmitted in the form of continuous beeps.

The result from the experiment has also proved that the Navbelt simulator ensures travel safety up to generally 0.8m/s in a very unfamiliar environment. The Navbelt uses some systems like modify sensor system which work on three operating modes offering a different level of assistance and consumer effort whereas the computer uses the stereophonic signals to give real-time representation of the current environment and then transmitting the result to the user but there are some changes before it comes to the user-handling that is needed to be done. Navbelt also has eight directly-pointing sensors but some additional sensors are needed to be developed with sensory capabilities. The other one is the implement positioning feedback system that has a GPS integration to receive feedback in the outdoors as GPS was first introduced in the 1990s but later on, in development, they proved to build another object as an obstacle to receive signals thus this device has some or the other kind of limitations. The last one incorporates head positioning system guides signal on auditory locations relating the user using the stereophonic headphones referring to these headphones they represent the direction concerning the user's head.

Thus when the head is turned the virtual direction changes and also due to the head position sensor the system detects moments relative to the body, so it can be concluded from above that implementation of this system in our Navbelt could be of greater use.

Iwan Ulrich et al. (2001) presents a new device termed GuideCane to help blind by navigating them through obstacles and hazards safely. As soon as the obstacle is detected by the ultrasonic sensors embedded machine within it decides the motion path guiding the GuideCane and the user about the environment around them. As

Guidecane is fully integrated system that makes use of all the components that are onboard being size and weight as key constraints in the whole mechatronic architecture but the design should be of the lighter weight so that user can carry it easily even though the public transport also the electronic and mechanical hardware both should work together for greater efficiency of the program. Being similar to the multi sonar array there is no need of scan being done by the user but there are some issues like detection of overhanging obstacles, upward-facing sonars that has not been addressed in this particular paper thus analyzing the situation and determining a suitable direction to avoid all the obstacles around the barrier is of much greater challenge. Thus taking a conclusion from this guide is only liable for local navigation missions as it only allows the user to work only on global navigation tasks.

Mounir Bousbia Salah et al. (2007) developed a navigation tool that helps in navigating the build people being based on the synthetic speech output micro-controller consisting of two vibrators, two ultrasonic sensors, mounted on the shoulder of the user. As the main device consists of the accelerometer, process microcontroller, footswitch, a hexadecimal keypad, and a power switch. The other subsystem contains a system fitted with ultrasonic sensors and wheels. As the individual move like a white cane, the style device notifies the obstacle by the center of the path as the wheels hit the ground the user can identify the ground conditions such as stairs and cavern. This detection system uses the ultrasonic signal of 40khz to acquire information and sending an ultrasound pulse.

Thus when the pulse is reflected by the path of the solid object the transmission time is usually calculated between the transmission and the source and this information is sent the user by which can locate the obstacle in his way. This same technique has also been used in the aircraft navigation system as well. Having a footswitch is the biggest advantage to the system as it provides great electronic travel help but there is also a disadvantage as the system detects the nearest barrier but cannot overcome the problem of perception but if we conclude as a whole the system can help the blind effectively.

Daniele Bernabei et al. (2011) have built and tested a Blind aid program that allows the user to experience the virtual environment. As the main objective is the assessment of modalities and navigation devices along with spatial-visual mapping for the blind people. The study that has included four participants (who are completely blind) shows the working of the system how the system has created a cognitive map utilizing virtual exploration instead of using virtual space. This paper has also delivered good results on the continuity and applications of this particular research.

The spatial awareness has also been applied through real-space orientation exercising in the corresponding real-space. The blind aid program also plays a good role in the implementation of newly blind O&M Recovery training & diagnostic tool to monitor the spatial participants. The blind Aid also supports the adventitiously blind on pre-exploration and the collection of spatial information as spatial knowledge on this can also be collected by the means of the internet which is pretty similar to the visual maps available on the net. Therefore, a spatial multimode device can help a large section of the population in gathering the spatial information for the navigation irrespective of the place where they want to bring it in use.

Ramiro Velázquez et al. (2010) provides a glimpse of the bio-medical applications in terms of wearable system which has a great role as assistive devices for the disable people having physical or sensory disabilities leading them a new way in their particular lives to understand the work that has been done on this in the past we need to look through the whole parameters of all wearable devices and systems which are built for the blind. The other point that matters is the cost so keeping the cost in mind these wearable devices are probably low cost and affordable for the blind built with off-the-shelf technology. Coming to the working, the Kinect c device which is coupled with the data takes the input from the depth maps which are produced from the same device, now as the accelerometer provides a 3D representation of the all the registered points to the user developing a time-critical framework to analyze the scene and made a classification on the base of the movable obstacles and then delivering feedback to the user.

So rather than using a camera to have a look on the user, we use the Kinect series camera which is mounted on the helmet of the user, which makes the Kinect in motion state rather than be stationary but if we see the past work stationary cameras has mostly brought to use for the gesture recognition. Kinect being very advantageous its resilience against all the odds like rapidly changing background does not bring much of the impact on the system, so even sometimes the background is in motion which happens in our case also the conventional camera which is also wearable does not give more difficulty in managing.

D. W. Schloerb et al. (2012) is partially based on the hypothesis of supplying appropriate information is both conceptual and perceptual through various channels assisting the blind people with anticipatory exploration. This research essentially, a new blind aid system has been designed and experimented having two goals which are, Firstly evaluation of (haptic and audio) for the navigation tools and secondly cognitive mapping. It is based on approximately ten people who are completely blind so instead of navigating in the real-space the verbal descriptions have brought up in use providing greater validity for the trade-offs. Further, this study has examined the individual in a volatile environment to apply spatial knowledge in real-space tasks.

Also, the application of a cognitive map in the real-space has a big impact on the orientation allowing the people to operate with each other or with the devices independently. For the implementation level, the designed system has proven its potential in mainly four aspects which include spatial behavior, spatial strategy and problem-solving. As the number of people whole are traveling irrespective of business or leisure are increasing day by day also the spatial knowledge is also increasing and for serving this purpose multimodality consisting of audio, visual and haptic can provide a piece of great information for the broader population.

Viltor Filipe et al. (2012) uses a system that assists the blind in navigating when exposed to indoor environments. The data is acquired using the Microsoft sensor that is mapped into the pattern representation and by using neural networks the system extracts the information and classifies it based on relevant features enabling the detection of the possible obstacles. As this Microsoft Kinect sensor is not pretty expensive and is available widely makes it a very good replacement to the other sensors that were brought to use previously. It also supports the feature set and can work in low-light environments not only this but this system also identifies the low possible imprints.

Obtaining information from the surrounding environment using artificial sensors, and acting accordingly, is becoming more common nowadays. The possibility of creating technology that simplifies the daily life of a person with special needs is easier today. Technologies that can analyze, in real-time, the surrounding environment and produce useful and interactive information are summed advantage. To conclude, there is also a neural network present that particularly classifies all the features by the use of varied distributed panels and therefore generalizing it to produce the outputs.

Alexy Bhowmick et al. (2014) has proposed a system that helps the blind people to travel as an electronic travel aid is a form of assistive technology having the purpose of enhancing mobility for the blind. It also helps a blind person with assistance about walking routes by using coded sounds to point out what decisions to make. While the development of other devices to aid visually impaired people in their everyday life has been increasing and in some cases, adequate solutions to providing sensory supplementation have been very effective in various manners but still, a lot of work is needed to be done. When it comes to outdoor environments there come a critical boundaries on which a much broader study has to be done as this navigation system provides an aid for the blind by gathering the data from the depth sensor even if the user is in the outdoor environment but With the advent of the possibilities of remote sensing in the form of ultrasound and radar more research effort was directed at the problems of remote sensing of the environment for visually impaired people.

The most significant hindrance associated with this is probably the design structure which causes some or other types of issues while providing the navigation aids, thus these issues are needed to be addressed from time to time to ensure the smooth functioning of the whole system. The technique on which the whole system is based on the consideration of the background noises or possibly use the self-designed clicks for avoiding obstacles.

Samleo L. Joseph et al. (2015) research provides the concepts of the technological idea of detecting obstacle detection and avoiding the blind navigation which is commonly brought up by the computer and then extends to the applications of machine learning to which is very fertile for such variety of researches. Because many systems have multiple skills like user-friendly or fully optimized they have gained a lot if we see them with the eye of the visual impair community. But if we go to the foundation of this particular paper for assisting navigation the wearable sensors have fostered a very situational awareness for the blind. The system also acquires a majority of the social messages to calculate the relevant features of the event and create an alert for the user. While the social semantics that has been captured requires various parameters for the reasoning and querying the landmark localization. One of the most challenging problems in the domain of autonomous aerial navigation is the designing of robust real-time obstacle detection and avoidance system. This problem is complex, especially for the micro and small aerial views, that is due to the presence of constraints. Therefore, using lightweight sensors (i.e., Digital camera) can be the best choice compared with other sensors; such as a laser. Even for real-time applications, different works are based on stereo cameras to obtain a 3D model of the obstacles or to estimate their depth.

The major restraint to this can be seen as the feasibility evaluation for our brought in use algorithm as sound localization can also even sometimes leads to many problems so real-time localization technologies have to be considered or brought up in the use.

João Guerreiro et al. (2017) investigation allows the virtual navigation in the field of data exploration to increase the spatial knowledge in consideration with the present environment, to avoid the fragmented problems of the real-world representing the sequential environment. Also, the results have shown very significant differences between the virtual Walk and virtual Leap building an accurate sequential representation. Many of the users we also able to locate the nearest approximate location or the POI in their nearest street blocks which delivers a great exposure to the real world. Participants in the particular research will also experience these interfaces, navigating around one or more of the textual, object, and hybrid worlds. The object world is an urbanized landscape which is having strong spatial and weak textual features. Similar to the noon landscape, this world has strong color and lighting cues. Also, the hybrid world is a bit similar, but with weaker spatial and stronger textual elements but we also know that the textual world has weak spatial and strong textual features. Many approaches enable the learning world to identify a route that is probably independent and contains subjective feedback which can simulate the navigation options.

The future works include the implementation of a larger amount of information or data without the overwhelming users but such knowledge or interaction is probably done in the simulation model which helps in understanding the time convey to the navigation options and provides adequate information to the real world.

Muhammad Nazruddin et al. (2018) uses the neurological and physical condition of the blind people and helping them walk through the busy roads and also even traveling to very new places. The main aim of this particular paper is to introduce a small-hand carrying system that can act as a great substitute for the walking stick. As in the past, many things like mobility can has been worked upon in which generally you sense the environment in which you are traveling and then you define the obstacle and try to avoid them which make the system difficult to understand to bring under the process so there has been developed from time to time which also is shown in this particular paper.

This device containing an ultrasonic sensor and optical sensors that detect the hindrance to a very significant level even though it is to a very far distance. The best part of using this particular sensor is the fact that they do not get affected by any of the obstacles which can be anything like dust or even any other unwanted material.

There are some shortcomings to this idea and implementation plus they cannot deal with the variation in temperature and also the calibration of these sensors is a big issue that directly and indirectly depends on their sensor models, also they have a very limited detection range. Thus various moderations are still needed to be done.

Muhammad Nazrul Islam et al. (2018) research paper comes up with a development of the walking assistants for the visually impaired people containing Arduino Digital magnetic compass which can provide the right direction to the user and simultaneously helps in obstacle detection. It has also a great feature of audio notification where the user will have all the recorded messages instead of even beeps. The other major benefit to this is also the fact that it has an emergency location tracking via the GSM module, these things are only possible because of the technological advancements over the years. Many researchers and scientists have been working for the betterment of the disabled or the blind people specifically by contributing in the technological sector, their contribution is remarkable. GSM is counted together with the other technologies as the evolution of wireless mobile telecommunication that provides high-speed circuit-switched data enabling the data rate for the universal telecommunication service.

Putting safety as their priority all the system are needed to be analyzed and then brought to use as vary components has their demerits and the user is completely relying on this particular system, being the main aim as independent mobility of the visually impaired people so we need to ensure that their traveling becomes easy and helps them in covering the distances without any hurdle.

Ramiro Velazquez (2010) study proposes a mobility aid for the blind people along with the wearing assistive so that all these devices can contribute together and help the blind people providing them an assisting aid, as the data shows that there are near about 130 million people in the world who are disabled to see so the contribution in building up of such technology is huge and also needed to be appreciated.

Problems that are particularly related to information transmission are concerned with reading, character recognition and rendering graphic information about 2D and 3D views but the most successful reading tool is the Braille dot code which was Introduced by Louis Braille in the 19th century, which has become a standard worldwide. Inventions have been done in the past which address the problems of character recognition and mostly pictorial representation consisting of tactile displays. They also permit character and graphic recognition by feeling a tactile version of them. With the revolution on the internet in the last years' many problems related to computer access for the blind arose but along with them, the solutions have also been designed such as voice synthesizers, screen magnifiers, and Braille output terminals. Voice synthesizers which are practically used for reading the computer screen, screen magnifiers enable the on-screen magnification for those with low-vision and the Braille output terminals are plugged to the computer so that information on the screen is displayed in Braille.

If we conclude as a whole there are several universal design concepts made for the acoustical/tactile based assistive devices have been presented as their guidelines add to stimulate both hearing and touch to obtain the best performance from these senses but there are still some considerations that must be taken into account while decision making.

Robert k. Kretschmann et al. (2018) reviewed a paper on safe local navigation for the visually impaired once which includes a haptic feedback device containing an array of lidars and the vibrotactile units, also hand free and wearable. This device generally consists of two systems i.e haptic strap and sensor belt, as the haptic belt helps in calculating the distance whereas the sensor calculates the time of flight distance. There have been several works in the past that have proposed the steps towards solution building to such a problem. A better example can be one study that is reported by the design of a handheld device consisting of two infrared sensors generally combined with audible and vibratory feedback. While users were not comfortable using the device in place of a white cane and the device was not successful in the prevention of collisions with the concerned environment. While in another study based on the same features and concept, a walker who was equipped with two laser scanners, computers, and a vibration motor in each handle enabled a user to navigate through an environment while holding on to the walker. We had also recently shown an alternative approach to the problem to develop a solution for this problem making use of a wearable depth camera which is also implemented with the algorithms in the work on a low-power vision processor helping in visual recognition.

The varied performance of the designed ALVU (Array of Lidars and Vibrotactile Units), both in the quantitative and qualitative results-driven, shows very much that it is possible to provide better solution for the sightless users which is also useful for navigation information without requiring any physical contact with the environment. The user can also perceive a piece of distance information from multiple directions at once without needing to probe them. We hope that this will increase the independence of the user while also allowing them to be discreet and functional in a world designed by sighted people.

Jinqiang Bai et al. (2018) reviewed a paper on virtual -blind road following based wearable navigation device specifically for the blind people in locating, way-finding, route following and also obstacle avoiding modules mainly which are the essential components in a navigation system, while it remains a challenging task to consider obstacle avoiding during route following, as the indoor environment is complex, changeable and possibly with dynamic objects. To develop a solution and address this issue, we propose a novel scheme for the utilization of a dynamic sub-goal selecting strategy to guide the users to their destination and then helping them by bypassing obstacles at the same time.

They also gives a detailed description of the hardware configuration of the proposed navigation device demonstrating the proposed blind navigation system, also there is a technique based on a QR code-based system which requires proximity recognition, nevertheless, the blind people have difficulty in approaching the QR code. However, also the dynamic obstacles and changes in the environment were not considered even in the past studies not able to help the blind in avoiding the dynamic obstacles.

This special paper concludes by presenting a very novel navigation device for visually impaired groups to provide help them reach the destination safely and efficiently even though the indoor environment.

2.3 Conclusion

Considering the decades of technological advancement, the solution to provide blind people has only been addressed by the minority and educational segments. The numerous feasible approach presented to fix this problem was by creating a portable device with a collection of sensors. The obvious sensor choices were Ultrasonic sensor, Optical Sensor and Laser Sensor, these sensors worked perfectly as intended to but failed to overall address the issue rather performed well at one specific task of the solution. The details of these challenging factors are compared directly with the current technological advancement to comprehend the diverged focus area and their limiting elements. The current trending approach using Microsoft Kinect depth camera has the potential to label most of the drawbacks. Also, not only navigation being a primary focus point, other areas such as using mobile phones and computers have advanced by enabling voice assistance, talk-back and screen magnification features. The basic requirements for these problems are developing a highly protable, low weight, a user-friendly and interactive device to function on all use-case scenarios just like a normal human encounter and finds solutions using cognitive thinking. Thus, similar to all the research, this paper will address a specific puzzle that was not viewed previous to navigate thought the curves on the road.

Chapter 3: Methodology

3.1 Introduction

The problem is discussed in three different aspects in Data Analytics. To undertake blind turning angle change in the curvy roads, a dataset of walking on road and accelerometer is applied to implement in the CNN Regression problem. Then, classification problem direction indication as left, right and forward applied on Forest Trails for Mobile Robots. Finally, Lane Detection pattern and their challenges in the current scenario.

3.2 Deep Learning Models Background

3.2.1 CNN Model

The method of machine pattern recognition has been revolutionized by CNNs over the decades. Many image pattern recognition tasks were conducted before the widespread adoption of CNNs using an initial stage of hand-crafted feature extraction accompanied by a classifier. The model's significant discovery is that ability at automatically training examples. When applied to image recognition functions, the CNN method becomes particularly powerful as the convolution process captures the 2D structure of images. Relatively few parameters need to be learned compared to the total number of operations by using the convolution kernels to scan an entire image.

Although CNN's with learned features have been widely used for more than twenty years, thanks to two important developments, their popularity has exploded in recent years. The first large labeled data sets are now widely available for testing and evaluation, such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Second, CNN learning algorithms are now being applied to massively parallel graphics processing units (GPUs), allowing learning and inference performance exponentially accelerated.

A common CNN model is composed of several kinds of layers such as the Convolutional layer, Pooling layer (downsampling), Fully connected input layer, Fully connected layer, Fully connected output layer in which each layer performs their task on image to result in very high accuracy. The topmost convolutional layer creates a feature map to predict the class probabilities for each feature by applying a filter that scans the whole image, a few pixels at a time. The second pooling layer also called downsampling scales down the amount of information the convolutional layer generated for each feature and maintains the most essential information (the process of the convolutional and pooling layers usually repeats several times). The third fully connected input layer flattens the outputs generated by the previous layers to turn them into a single vector that can be used as an input for the next layer. The forth fully connected layer applies the weights over the input generated by the feature analysis to predict an accurate label. The last and the final fully connected output layer generated the closing probabilities to determine a class for the image.

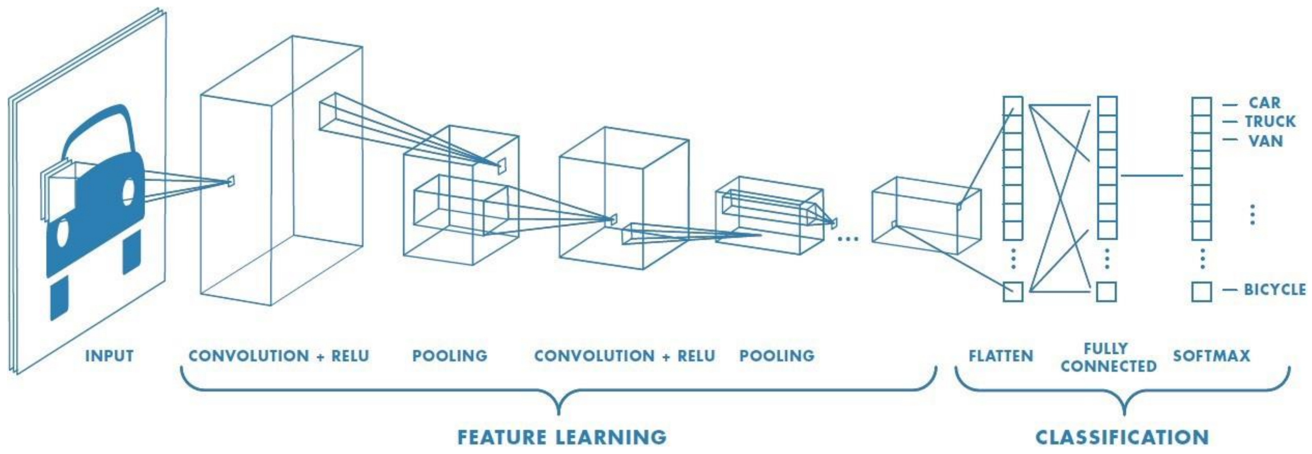

Figure 1: CNN Architecture

3.2.2 AlexNet CNN Model

Convolutional Neural Networks (CNNs) have always been the go-to platform for object recognition and they are strong structures that are easy to control and simpler to learn. When used on millions of images they don't suffer overfitting at a disturbing rate. Its efficiency is approximately the same as typical neural feedforward networks of the same scale. The difficulty in this network is that high-resolution images are difficult to apply. An advancement had to be tailored for GPUs on the ImageNet scale and to minimize training times while improving performance.

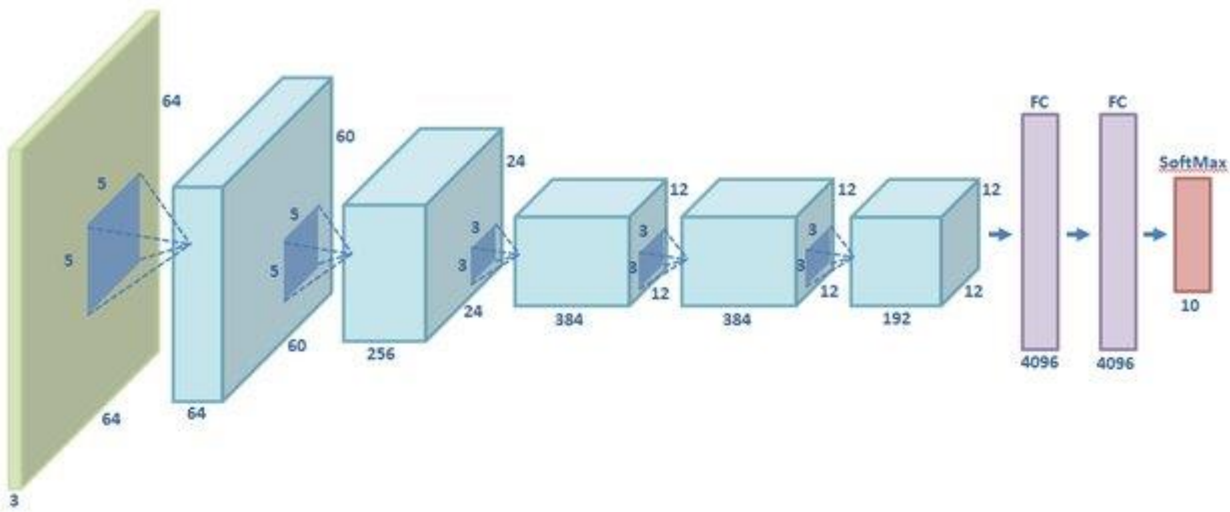

Figure 2: AlexNet Architecture

The framework consists of eight layers of five convolutional layers and three fully connected layers, and new approaches to convolutional neural networks are being used for a few more capabilities. Instead of the tanh function, which was then normal, a Rectified Linear Units (ReLU) non-linearity applied. The value of ReLU is in training time; a CNN using ReLU was able to reach an error of 25 percent six times faster on the CIFAR-10 dataset than a CNN using tanh. It uses multiple GPUs, with 3 gigabytes of memory rolling around. By placing half of the model's neurons on one GPU and the other half on another GPU, AlexNet allows multi-GPU training. This not only means that it is possible to train a larger model, but it also reduces the training time. An overlapping pooling, a typical "pool" CNN production of adjacent neuron classes without overlap.

Nevertheless, they saw a decrease in error by about 0.5 percent when the developers implemented overlap and noticed that models with overlapping pooling usually find it harder to overfit. Data Augmentation is used to render their data more diverse as a brand-preserving process. In particular, image translations and horizontal reflections were generated, which increased the training set by a factor of 2048. They also carried out Principle Component Analysis (PCA) on the RGB pixel values to improve the RGB channel intensities, which decreased the top-1 error rate by more than 1%. The dropout strategy consists of a fixed likelihood (e.g. 50 percent) of "turning off" neurons. It implies that each iteration uses a different set of parameters of the model, which allows each neuron to have more stable properties that can be used with other random neurons. Dropout, however, often raises the training time needed for refinement of the model.

3.2.3 VGGNet CNN Model

VGGNet is a Convolution Neural Net (CNN) architecture that was used to win the ImageNet Large Scale Visual Recognition Challenge (Imagenet) competition in 2014. It is considered to be one of the excellent vision model architecture to date. The most unique thing about the VGG-16 version is that instead of having a large number of hyper-parameter they focused on having convolution layers of 3x3 filter with a stride 1 and always used the same padding and max pool layer of 2x2 filter of stride 2. It follows an arrangement of convolution and max pool layers consistently throughout the whole architecture. In the end, it has 2 FC(fully connected layers) followed by a softmax for output. The 16 in VGG16 refers to it has 16 layers that have weights. This network is a pretty large network and it has about 138 million (approx) parameters.

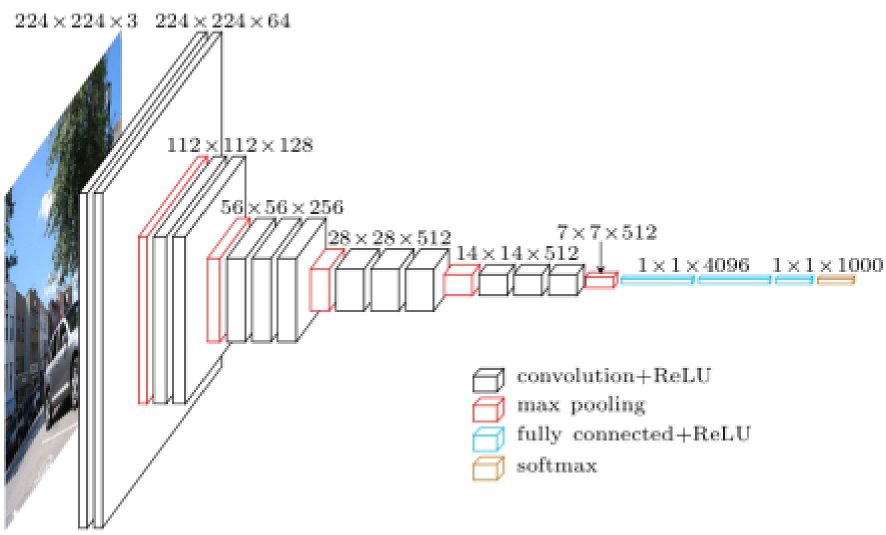

Figure 3: VGG-16 Network Architecture

3.2.4 NVIDIA's CNN Self-Driving Car Architecture

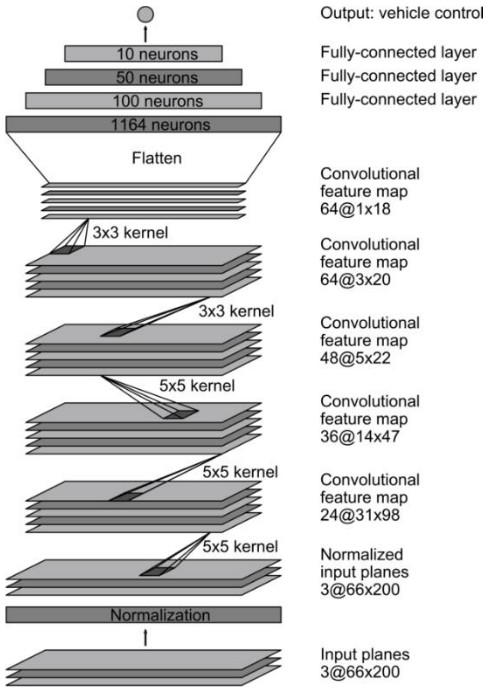

Figure 4: Nvidia CNN Network Architecture

In the custom NVIDIA model, the first layer of the network performs image normalization. The normalizer is hard-coded and in the learning process, the values are not adjusted. Performing network normalization enables the normalization scheme with the network architecture to be altered and optimized by GPU computing.

The convolution layers are configured for the extraction of features and are empirically chosen through a series of experiments that change the configuration of layers. In the first the convolution layer, the use of strided convolutions with a 2*2 stride and a 5*5 kernel, and a non-strided convolution with a 3*3 kernel scale in the last two convolution layers.

The five convolutional layers with three fully connected layers are matched, resulting in the final output control factor being the reverse-turning-radius. The fully connected layers are designed to function as a direction angle, but noticed that a clean break between these sections of the network mainly operates as a feature extractor and which serve as the controller is not feasible through training the system end-to-end.

3.3 Navigation using Classification & Regression Analysis

3.3.1 Data Source

The problem is split into two groups one uses regression technique with self-recorded footpath data along with accelerometer to identify turning angle and another is a classifier to labeling directions namely left, right and forward on "Visual Perception of Forest Trails for Mobile Robots" dataset.

3.3.1.1 Classification

The data has been taken from the website people.idsia.ch. The data is a collection of images which have been collected from a set of people who worked on a project on visually percieving forest or mountain trails from a images captured by a robot striding on the trail programmed. The images in the dataset were shot using 3 Bluefox cameras, which is fixed rigidly on the helm. The dataset is 1.2 GB in size and the images are stored in the folder “videos” and further in three subfolders which are “lc”, “rc” and “sc”. Each image is in the jpg format. Images for paths which are for routes on the left of the robot are stored in the lc folder, routes on the right are in the rc folder and routes which are continuing on the straight are in the sc folder.

3.3.1.2 Regression

The core concept of this paper is inspired by an autonomous car containing a steering angle and a front-camera record. Similarly, it is necessary to build a source that is simple yet powerful to run both machine learning and deep learning models. To capture the data, a bike-mounted with two smartphones on the handlebar is employed as in the above figure. The model of the two phones is Oneplus 6T and Nexus 5. Using these two smartphones, the data is collected, Oneplus 6T is used to record the video which is kept in a vertical angle and the Nexus 5 to capture the reading of the accelerometer. The recorded data from the smartphones contains the video of the footpath and accelerometer reading X, Y, and Z coordinates along with the time.

3.3.2 Data Understanding & Analysis

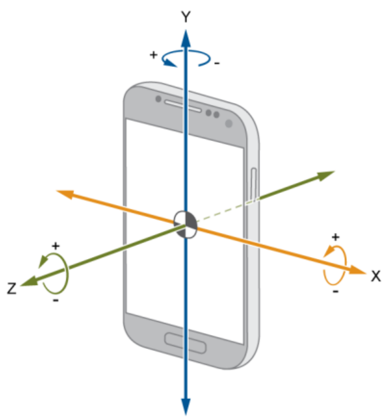

Figure 5: Phone Accelerometer Axis

An accelerometer in mobile phones is used to detect the orientation of the phone in x, y, and z coordinates. In the use case, y coordinate is used to analyze the angle of the turn taken by the handlebar of the bike which metaphorically symbolizes the human turning radius in lateral position. The y cooridinate value ranges from -0.3 to +0.3 The video is recorded in Tymon Park, Dublin for about 12:29 minutes with 30 frames per second and the frame width and height are 1280 by 720 respectively and overall the size the video is 1.23 Gigabytes.

An OpenCV open-source library for computer vision is chosen to transform each frame of a video to an image. The video is a 720p high-definition of 1280 pixels by 720 lines. To reduce computational time, the transformed images are scaled down on both width and height to 60 percent and decreased the overall quality by 30 percent. Each file name begins with "image_<count>.jpg" and this information is stored in a comma-separated file (CSV) also with the turning angle. Finally, before applying a model, it a practice to split into training data and test data. The training set contains a known output and the model learns on this data to be generalized to other data later on. The test dataset (or subset) to test our model's prediction on this subset.

3.3.3 Exploratory Data Analysis

It is a data analysis methodology that utilizes a range of (mostly graphical) methods to optimize the visibility into a data set, detect underlying structure, isolate relevant variables, detect outliers and exceptions, check underlying assumptions, build parsimonious models and evaluate optimal factor settings.

Recent advances in deep learning algorithms are primarily due to the amount and variety of data collected in recent years. Data augmentation is a technique that encourages practitioners to significantly increase the variety of data available for training models without actually collecting new data. The basic methods such as cropping, padding and horizontal flips with massive neural networks. Moreover, the majority of methods used in neural networks training utilizes only basic types of augmentation. While the in-depth analysis was carried out on neural network architectures, less emphasis was focused on finding specific forms of data augmentation and data augmentation policies collecting data invariances. The techniques include affine transformations, perspective transformations, contrast changes, Gaussian noise, dropout of regions, hue/saturation changes, cropping/padding, blurring are used to produce variations in the image. The functionality to augment images with segmentation masks, bounding boxes, key points, and heatmaps methods act as a key for transformation.

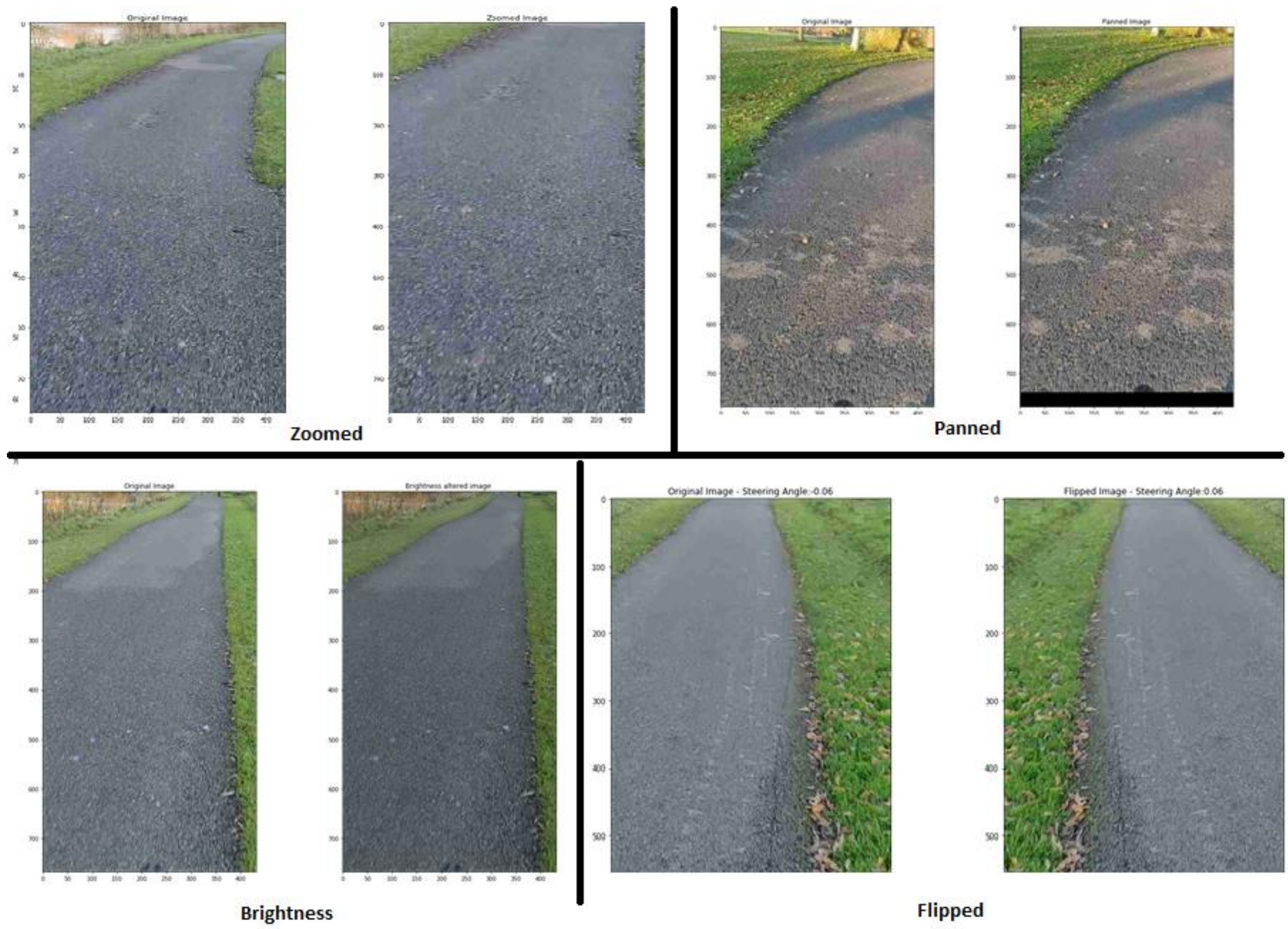

Figure 6: Augmentation

The above images show a particular image with a relating transformation performed on it. This has been done to understand the various aspects of a particular image. Each image can be understood better by implementing operations such as panning an image to discern additional details by removing a small area that is well defined. This helps in focusing more on newer details of the image which would be ignored otherwise as it contains lesser attributes that could be helpful for prediction. Flipping which results in the mirror of the same image. This gives insights into the image from a different dimension thereby unraveling more attributes that would be hidden. Also, altering the brightness where the brightness has been lowered thereby sharpening the overall quality of the image to make the algorithm accept more aspects of the image.

3.3.4 Results

The experimental results have proven the findings that concluded that NVIDIA and VGGNet were the best performing classification models with 56% and 55% accuracy respectively. AlexNet produced the least accuracy of 46% among the three classification models. A custom CNN produced the highest accuracy of 82% among the two regression models and can be considered as our best regression model. It was closely followed by NVIDIA with an overall accuracy of 76%. A good value for Accuracy has been achieved from the simple CNN model.

3.4 Lane Detection Analysis

3.4.1 Introduction

Line detection is as important as edge detection in lane detection. This typically includes two models that include a feather-based method and model-based methods. Some use modified Hough transform to remove lane profile segments and use the clustering algorithm DBSCAN (Density-based spatial clustering of applications with noise). Others use progressive probabilistic Hough transform coupled with the technology of maximum stable extreme area (MSER) to define and detect lane lines and use the Kalman filter to achieve continuous tracking. Nevertheless, the algorithm does not work well at night. However, in this case Hough Transform is used to detect the lane.

3.4.2 Process of Edge Detection

The goal of Edge Detection is to identify the boundaries of objects within images. In essence, this technique is used to find the regions in an image where there is a sharp change in the intensity in adjacent pixels. It is important to recognize that an image can be arranged as a matrix which is an array of pixels that contains the light intensity at some location in the image. The intensity of Each pixel denoted by a numeric value that ranges from 0 to 255 and the intensity value of zero indicates no intensity as completely black and 255 represents maximum intensity gradient is that bright over a series of pixels. A strong gradient indicates a steep change whereas a small gradient represents a shallow change. This helps us identify edges in our image since an edge is defined by the difference in intensity values in adjacent pixels. The sharp change in intensity and a rapid change in brightness results in a strong gradient that appears in a bright pixel in the gradient image by tracing out all of these pixels we obtain the edges.

The first in the process to convert an image to grayscale. The approach behind it is in a greyscale image only one channel on each pixel with only one intensity value ranging from 0 to 255 whereas a three-dimensional colored image that will include red, green and blue channels and in each pixel a combination of three intensity values are present. The point is by using a grayscale image processing a single channel is faster than processing a three-channel color image and less computationally intensive

The second step is to reduce the noise and accurately catch as many edges in the image as possible it is necessary to include a filtering mechanism on any noisy otherwise, false edges detection will ultimately affect edge detection. It is imperative to filter it out and thus smoothen the image. Filtering out an image noise smoothing will be performed with a Gaussian filter $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x2}{2\sigma^2}}$. To understand the concept of a Gaussian filter, recall that an image is stored as a collection of discrete pixels. Each of the pixels for a greyscale image is represented by a single number that describes the brightness of the pixel. Typically to modify the value of a pixel with the average value of the pixel intensities around it. Averaging out the pixels in the image to reduce noise will be done with the kernel. Essentially this kernel normally distributed numbers is run across our entire image and sets each pixel about equal to the weighted average of its neighboring pixels thus smoothing our image.

The stage is set to apply the Canny method to identify edges in our image. Recollect that an edge corresponds to a region in an image where there is a sharp change in intensity or a sharp change in color between adjacent pixels in the image. The change in brightness over a series of pixels is the gradient. A strong gradient indicates a steep change whereas a small gradient a shallow change. The basis of an image is itself composed of pixels that can, therefore, be read as a matrix. An array of pixel intensities to compute the gradients in an image, it must recognize the representation of the image in a two-dimensional coordinate space X and Y. The x-axis traverses the image's width and the y-axis goes along the image's heights. Also, width and height representing the number of columns and rows in the image respectively. The product of both width and height yields the total number of pixels in your image. The purpose being is not only to look at an image as an array but also as a continuous function of x and y $\boldsymbol{f(x,y)}$. The Canny function will perform a derivative on the image function in both x and y directions, thereby measuring the change in intensity toward adjacent pixels. It computes the gradient in all directions of our blurred image and it is then going to trace the strongest gradients as a series of white pixels. But notice these two arguments low threshold and high threshold. While this actually allows us to isolate the adjacent pixels that follow the strongest gradients. Firstly, if the gradient is larger than the upper threshold then it is accepted as an edge pixel and then if it is below the lower threshold it is rejected. Finally, if the gradient is between the thresholds then it will be accepted only if it is connected to a strong edge. To detect a straight line in an image, a Hough transform technique will be implemented. A straight line is represented by the equation $\boldsymbol{y = mx + b}$. The change in y over the change in x which evaluates y-intercept and slope of this entire line can be plotted as a single point in Hough space. Split the Hough space into a grid on each been inside of the grid corresponding to a slope and y-intercept value of a candidate line. If the points of intersection are inside of a single bin. For every point of intersection, function cast's a vote inside of the bin and the bin with the maximum number of votes will be the line. This line is drawn since it has voted as the line of best fit in describing our data.

3.4.3 Report

Figure 7: Canny Edge Detector

The picture on the left shows the black and grey image of a path and then the corresponding image next to it after the Canny edge detector path finding algorithm has been applied to it. This technique has been applied using three parameters which are Sigma, Low threshold and High threshold. The technique works by first performing Gaiussian smoothing, then applying gradient and finally non-maximum suppression. The value of Sigma has been set to 10, Low threshold value is 28% and High threshold value is 42%. The resulting image after the applying the Canny edge detector, the path of the road is accurately extracted. The second image is of a pavement. The structure of the footpath is not entirely accurate, therefore the path detection algorithm was not able to correctly perform as in the previous illustration.

Chapter 4: Future

4.1 Proposal

Autonomous blind walk on a footpath in a virtual environment production is a less-studied problem. The Sensorimotor motor role remains a major obstacle for machine learning and Robotics. The densely populated urban area navigation becomes a key challenge to simulate and train. Due to complicated multi-agent interactions at traffic intersections and the need to monitor and react to the activity of tens or hundreds of other agents that may be in view at any given time. Legal traffic laws that identify road signs, street lights, and road signs and discriminate between multiple types of other vehicles, people, dogs, and movable and stable objects; a long tail of unusual road construction; a child on the road; an incident ahead; a delinquent car barricade on the wrong side.

Autonomous walk work is hindered by infrastructure costs and the real world's logistical challenges in training and testing technologies. On the fly requirement of sensorimotor control data in the physical world for walking with types of equipment are beyond the scope of the most research analysis. Training and validation of driving strategies in simulation is an alternative option. As some events are too risky to be performed in the physical world (e.g., a kid running on the path ahead of the car), it is also important for device testing. Since the early days of autonomous driving research, simulation has been used to train driving models.

Chapter 5: Conclusion

This paper presents a novel navigation challenges in implementing the smartphone assist aid for visually impaired groups to help them reach the destination safely and efficiently in an outdoor environment. The problem in the Data Analytics perspective is categorized into regression and classification. Regression for determining the turning curve and classification for the direction indicator. In both cases, a very popular Deep Learning CNN model is implemented to find the results. To predict the turning deviation a custom NVIDIA, VGGNet, and AlexNet models are applied and in return, NVIDIA and VGGNet performed better. On the other hand, in classification problems for guiding left, right and forward paths the standard NVIDIA performed did not perform well compared to simple CNN models. Finally, a Lane Detection problem is elaborated to recognize the challenges associated with it and also to resolve an issue linked to payment. Experimental results fairly performed in some cases. The reason for these downward results is a weak dataset, lack of simulations and footpath. Thus, visually impaired people walk from one place to another is still remains an unsolved puzzle.

References

Bernabei, D., Ganovelli, F., Benedetto, M., Dellepiane, M. and Scopigno, R. (2011). A Low-Cost Time-Critical Obstacle Avoidance System for the Visually Impaired. *INTERNATIONAL CONFERENCE ON INDOOR POSITIONING AND INDOOR NAVIGATION*.

Filipe, V., Fernandes, F., Fernandes, H., Sousa, A., Paredes, H. and Barroso, J. (2012). Blind Navigation Support System based on Microsoft Kinect. *Procedia Computer Science*, 14, pp.94-101.

Guerreiro, J., Ahmetovic, D., Kitani, K. and Asakawa, C. (2017). Virtual Navigation for Blind People: Building Sequential Representations of the Real-World. *Research Gate*.

IntelliNavi, A., Prakash, S., Bhagat, R., Prasad, V. and Hazarika, S. (2014). IntelliNavi : Navigation for Blind Based on Kinect and Machine Learning. *Research Gate*.

Joseph, S., Xiao, J., Zhang, X., Chawda, B., Narang, K., Rajput, N., Mehta, S. and Subramaniam, L. (2015). Being Aware of the World: Toward Using Social Media to Support the Blind With Navigation. *IEEE Transactions on Human-Machine Systems*, 45(3), pp.399-405.

Kumari, P. and ., .. (2018). Automated Human Identification and Obstacle Avoidance for Visually Impaired. *International Journal of Engineering & Technology*, 7(3.6), p.9.

Lahav, O., Schloerb, D., Kumar, S. and Srinivasan, M. (2012). A virtual environment for people who are blind – a usability study. *Journal of Assistive Technologies*, 6(1), pp.38-52.

Shoval, S., Borenstein, J. and Koren, Y. (1998). The NavBelt-a computerized travel aid for the blind based on mobile robotics technology. *IEEE Transactions on Biomedical Engineering*, 45(11), pp.1376-1386.

Ulrich, I. and Borenstein, J. (2001). The GuideCane-applying mobile robot technologies to assist the visually impaired. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 31(2), pp.131-136.

Velázquez, R. (2011). Wearable Assistive Devices for the Blind. *Research Gate*.

Bojarski, M., Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J. and Zieba, K. (2016). End to End Learning for Self-Driving Cars. *ArXiv*.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A. and Koltun, V. (2017). CARLA: An Open Urban Driving Simulator. ArXiv.

Seita, D. (2020). *1000x Faster Data Augmentation*. [online] The Berkeley Artificial Intelligence Research Blog. Available at: https://bair.berkeley.edu/blog/2019/06/07/data_aug.

Abeythilake, U. (2018). *Gyroscope and Accelerometer with ionic 3*. [online] Simple Activity. Available at: https://simpleactivity435203168.wordpress.com/2018/06/28/gyroscope-and-accelerometer-with-ionic-3/.

Llamas, J., M. Lerones, P., Medina, R., Zalama, E. and Gómez-García-Bermejo, J. (2017). *Classification of Architectural Heritage Images Using Deep Learning Techniques*.

Seita, D. (2019). *1000x Faster Data Augmentation*. [online] The Berkeley Artificial Intelligence Research Blog. Available at: https://bair.berkeley.edu/blog/2019/06/07/data_aug.

Thakur, R. (2019). *Step by step VGG16 implementation in Keras for beginners*. [online] Medium. Available at: https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c.

People.idsia.ch. (2017). [online] Available at: *http://people.idsia.ch/~guzzi/DataSet.html*